Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 47 No. 03 (2026)

Hybrid Tumor Classification Using Vision Transformers and PSO-Optimized Feature Selection with XGBoost

Seema Afreen Khan[1], Dr.NagaMani Chippada[2]
[1,2]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Green Fields, Guntur, AP, India.

Abstract

To enhance patient outcomes and efficiently direct treatment, it is critical to quickly and precisely classify brain lesions in magnetic resonance imaging (MRI). Our hybrid classification system combines an XGBoost classifier, Principal Component Analysis (PCA)-driven dimensionality reduction, Particle Swarm Optimization (PSO)-based feature selection, and Vision Transformer (ViT)-based feature extraction in a way that complements each other. We evaluate our approach using the Brain Tumor Classification (MRI) dataset from Kaggle, which consists of 3,264 T1-weighted images evenly distributed across four categories: healthy, meningioma, pituitary tumor, and glioma. To preserve 95% of the variance, the ViT encoder compresses the large amount of global context data it receives using PCA. By selecting the most valuable components, PSO further enhances this representation and reduces the feature count by over 70%. The final classification accuracy of 97.8%, 98% precision and recall, and 0.99 area under the ROC curve (AUC) are achieved by XGBoost using the optimized feature subset. With a mean accuracy of 96.5% (±1.0%) and a mean AUC of 0.985 (±0.005), five-fold cross-validation demonstrates the stability of the model. In addition to being more effective and simpler to comprehend, our pipeline performs on par with or better than the top CNN and transformer models. This study demonstrates that there is a powerful, scalable method for automatically diagnosing brain tumors using sophisticated deep learning architectures and evolutionary optimization. Other medical imaging applications might also make use of this technique.

Keywords: VisionTransformer,Principal Component Analysis ,Particle Swarm Optimization, XGBoost ,Data Augmentation (RandomRotation, Flips, Zoom, Shift),Stratified Shuffle Split,Brain Tumor MRI Classification.

Introduction:

Background and Clinical Relevance

In order to facilitate effective treatment, brain tumors necessitate a timely and precise diagnosis, as they are life-threatening neoplasms. Early diagnosis and accurate subtype classification (for example, glioma, meningioma, pituitary, or normal) are for making predictions and arranging treatment. For instance, CBTRUS says that more than 700,000 persons in the US [1-3] have a primary brain tumor and that roughly 84,000 new cases are found each year. Magnetic resonance imaging (MRI) is the usual noninvasive way to see brain tumors, but it takes a long time to interpret by hand and can vary from person to person. Automated classification of brain MRIs can help radiologists by quickly finding tumors and their types, which could lead to better patient outcomes and more efficient workflows. Deep learning has become a significant technique for solving this problem by advances in deep learning.

Learning for Medical Imaging

For a long time, image-based medical diagnosis has been dominated by convolutional neural networks, or CNNs. Several CNN-based models have demonstrated high accuracy in the classification of brain tumors. To differentiate gliomas, meningiomas, and pituitary tumors from MRI [1, 2, 3, 4, 6, 11], for example, Deepak and Ameer (2019) used a pre-trained GoogLeNet and achieved an accuracy of roughly 98% [1]. Similarly, Francisco et al. (year) reported a multi-pathway CNN with an accuracy of about 97.3% [2] in classifying three types of tumors. Using multi-class MRI datasets [3.4, 11], transfer learning using architectures such as ResNet, VGG, and EfficientNet has also shown accuracies in the mid-90s.

More recently, attention-augmented CNNs (like CBAM) and cascaded pipelines have made things much more resilient, especially on scans that are noisy or have low contrast [2, 8, 15]. Generative methods for data

augmentation, like GAN-based two-step pipelines, have been shown to enhance the performance of classifiers by increasing the size of small MRI datasets.

Transformer-based architectures have demonstrated significant promise in medical imaging and computer vision in more recent times. Self-attention is used by Vision Transformers (ViT) to identify heterogeneous tumor patterns by capturing long-range dependencies across an image. Using a dataset of 5712 images of brain tumors, Carlsen et al. (2023) improved a ViT on contrast-enhanced MRI and reported 98.13% classification accuracy. Reddy et al. (2024) evaluated a number of refined ViT variants on a sizable brain MRI collection (7,023 images, four classes) and found that the optimal ViT model could achieve up to 98.7% accuracy. These findings show that ViT can learn rich, discriminative features from brain MRI, frequently outperforming traditional CNNs on the same tasks. Ensemble and hybrid approaches have been investigated in addition to deep networks; for instance, combining CNN feature extractors with machine learning classifiers has enhanced classification performance compared to using just one model.

Limitations of Existing Methods

Current methods have significant drawbacks, despite their high reported accuracy. Generally speaking, deep CNNs and ViTs need a lot of computation and large labeled datasets; on smaller or unbalanced medical image sets, their performance may suffer. Instead of concentrating on general multi-class classification, many studies concentrate on binary tumor-versus-normal classification or on particular tumor subtypes. Some earlier research, for instance, only differentiates between tumor and no tumor, neglecting the crucial task of subtyping into gliomas, meningiomas, pituitary, etc. Overfitting is a risk when training sophisticated CNNs on sparse MRI data, and deep models' "black-box" nature can cause issues in clinical settings. Additionally, feature reduction is not explicitly carried out by CNN-based pipelines, which can leave behind redundant or noisy data that impedes learning. On the other hand, when given optimized feature subsets, traditional machine learning classifiers can be more robust and frequently perform well with smaller feature sets. Therefore, increase generalizability and interpretability, hybrid approaches that combine deep feature extraction with principled feature selection and effective classifiers are required[5,13,14,16,18,19].

Motivation for a Hybrid ViT–PSO–XGBoost Approach

We suggest a hybrid pipeline that combines the advantages of ensemble classifiers, dimensionality reduction, evolutionary feature selection, and Vision Transformers. The ViT component encodes global image context into high-dimensional embeddings, acting as a potent feature extractor. These unprocessed features, however, may be noisy and redundant. tackle this, we first compress the ViT output into a lower-dimensional space using Principal Component Analysis (PCA), keeping the components that capture the most variance. PCA is a traditional method that projects data onto orthogonal axes alleviate the curse of dimensionality. Then, building on PCA, we choose an ideal subset of these elements using Particle Swarm Optimization (PSO). A bioinspired metaheuristic, PSO efficiently looks for instructive patterns in the feature space. According to earlier research, integrating PCA and PSO on medical image features can greatly improve classification. For example, Alnashwan et al. (2023) found that selecting only the most salient components increased accuracy. To further reduce feature redundancy in our framework, PSO iteratively finds the most discriminative PCA features for classes of brain tumors.

Finally, we use XGBoost, a group of gradient-boosted decision trees, to classify the selected features. People know that XGBoost works well with data that has dimensions and is resistant to overfitting. XGBoost has often done better than other classifiers on medical imaging tasks in studies that compare them. For example, one review found that XGBoost had the highest accuracy (about 92%) of all the classifiers on a standard brain MRI benchmark. PCA+PSO gives XGBoost a small, optimized set of features that works well in our hybrid model. The ViT encoder captures complex visual patterns, the PCA reduces noise, the PSO focuses on the most useful features, and the XGBoost classifier is powerful and easy to understand. Together, these modules improve performance. This combination should lead to more accurate and dependable tumor classification than any one method alone[7,12,17].

Dataset Description

We test our work on the Brain Tumor Classification (MRI) dataset from Kaggle (Sartaj Bhuvaji et al., 2020). About 3,264 T1-weighted brain MRI scans in four different categories—glioma tumor, meningioma tumor, pituitary tumor, and no tumor (healthy)—are included in this dataset. This "BT-large-4c" dataset has been segmented using an 80/20 train-test ratio in previous work. Each class is represented in each split, which yields approximately 2,611 training images and 653 testing images. Due to the diverse sizes and origins of the dataset photos, standard preprocessing techniques such as resizing, normalization, and data augmentation are employed. Every class has clinical significance. For instance, pituitary tumors and the normal class round out the range of

disorders seen in the dataset, while gliomas and meningiomas represent common primary brain tumors with particular treatment regimens. Our analysis aligns with recent literature and allows direct comparison to state-of-the-art methods by utilizing this popular dataset with numerous classes.

Summary of Proposed Method and Key Results

Each MRI is handled by our hybrid framework in this way: First, using patch embedding and multi-head self-attention, the Vision Transformer (ViT) converts the input image into a collection of deep feature vectors. To extract principal components that capture 95–99% of the variation, PCA is then applied to these vectors. The PCA output is then subjected to PSO, which selects the best components that are most predictive of tumor class. In order to predict the tumor category, the chosen features are then fed into an XGBoost classifier. Figure 1 illustrates this pipeline (ViT encoding ➔ PCA reduction ➔ PSO selection ➔ XGBoost classification).

The accuracy and discrimination of the proposed model are good. We achieve an overall accuracy of 97.8%, precision and recall of approximately 98%, and an area under the ROC curve (AUC) of 0.99 on the Kaggle dataset. These findings are in good agreement with previous research: optimized ViT models have achieved ≈98–99% accuracy on comparable multi-class MRI tasks. Additionally, our performance is comparable to that of other hybrid approaches; for instance, recent research that combined enhanced classifiers with deep features has achieved accuracies in the mid-90s. By concentrating on strong traits, PCA+PSO successfully managed possible class imbalance, as evidenced by the breakdown of our results, which shows balanced class performance (no single class dominating).

Key components and their impact:

- The proposed model has discrimination and accuracy. On the Kaggle dataset, we obtain an area under the ROC curve (AUC) of 0.99, an overall accuracy of 97.8%, and precision and recall of approximately 98%. These results are consistent with previous studies, which found that optimized ViT models achieved 98–99% accuracy on similar multi-class MRI tasks. Furthermore, our results are comparable to those of earlier hybrid approaches; for example, recent studies that combine deep features and improved classifiers have achieved mid-90s accuracies. The breakdown of our results shows balanced class performance (no single class dominating), indicating that PCA+PSO successfully addressed probable class imbalance by focusing on strong features.
- Focus, providing a detailed portrayal. The ViT features serve as a comprehensive classification foundation.
- **PCA:** Reduces computational cost and eliminates noise by compressing the high-dimensional ViT characteristics into a lower-dimensional subspace. PCA avoids overfitting and speeds up additional optimization by preserving principal components (e.g., 95% variance).
- **Particle Swarm Optimization (PSO):** The most instructive subset of PCA components. PSO improves feature selections, giving the classifier more discriminative inputs. Previous research has demonstrated that PCA+PSO selection provides better classification performance than using PCA alone or all features.
- **XGBoost:**smaller feature set to learn decision trees. XGBoost effectively manages feature heterogeneity and resists overfitting as an ensemble method. In our tests, XGBoost successfully used the enhanced features to distinguish between tumor classes with remarkable recall and precision.

  In conclusion, this hybrid ViT–PCA–PSO–XGBoost method uses standard machine learning for feature classification and refinement and deep learning for feature extraction. With precision ≈98%, recall ≈98%, and AUC ≈0.99, the results that each module contributes to the final accuracy of 97.8%. These metrics suggest that our approach performs on par with or better than the most advanced models for classifying brain tumors from MRI images. The suggested system offers reliable, high-fidelity classification suitable for clinical datasets by combining intelligent selection, boosting, and global feature learning.

Literature review:

Ali Akbar Siddique et al. (2023) introduced the PSO-INCEPT framework, integrating InceptionV3, Particle Swarm Optimization (PSO), and the ADAM optimizer for brain tumor classification using a Kaggle dataset with four classifications. The model was very accurate, getting 99.33% of the answers right. But because it depends on complicated optimization methods, it needs a lot of computing power, which could make it to use in places with limited resources.

B. Shyamala et al. (2023) utilized a regression neural network with Relief and PSO-based feature selection for brain tumor analysis using the Figshare dataset. Their method showed promise, but it relied significantly on

MATLAB and didn't compare itself to current deep learning models, which made it less useful in modern AI-driven medical imaging processes.

Bach Hoai Nguyen et al. (2022) focused on PSO-based feature selection across three real-world datasets. While their work contributes greatly to the optimization of input features, it lacks complete classification performance measurements and does not give an end-to-end deep learning pipeline, making it challenging to evaluate its full clinical relevance.

Changhee Han et al. (2023) applied a mix of GAN-based data augmentation (PGGAN, MUNIT, SimGAN) and CNN for brain tumor identification. The model obtained 97.48% accuracy, exhibiting strong performance. However, the technology is computationally demanding and lacks significant clinical validation, which may impede its practical application in healthcare settings.

Hasnain Ali Shah et al. (2023) fine-tuned EfficientNet-B0 for brain tumor categorization. Achieving 98.87% accuracy, the model displayed strong performance. However, the study lacked thorough details on the dataset used and depended significantly on data augmentation techniques, raising doubts about the generalizability of the conclusions.

K. V. Shiny et.al (2023) introduced an HBA-optimized U-Net architecture for brain tumor segmentation, obtaining an accuracy of 94.9%. Although the model performed relatively well, it did not outperform state-of-the-art methodologies and lacked comparative evaluation, limiting its potential for real-world implementation.

Kusum Lata et al. (2023) created an ensemble model incorporating ResNet-50, Inception V3, and VGG-16, obtaining 99.92% accuracy. Despite this, the combination of privacy protections with ensemble models increases system complexity and slows down processing, impacting its efficiency in real-time applications.

Mahnoor Ali et al. (2023) suggested an ensemble 3D CNN and U-Net framework utilizing the BraTS-19 dataset, concentrating on brain tumor segmentation. The model achieved a Dice score of 0.81. While effective in segmentation, the study minimal information on classification metrics, restricting its applicability for diagnostic classification.

Malliga Subramanian et al. (2023) studied MobileNet, VGGNet, and DenseNet paired with Learning without Forgetting (LwF) and Bayesian optimization for interpreting CT/MRI images of eight cancer kinds. Although promising, the study lacked tumor-specific insights and focused more widely on multi-cancer detection, limiting its specificity in brain tumor diagnosis.

Manoj Ishi et al. (2023) applied GLCM for feature extraction, PSO for selection, and CNN for classification on a dataset from the Harvard repository. Achieving 96.71% accuracy, the model highlighted texture-based elements. However, its reliance on simple preparation techniques may limit performance on more complicated or noisy datasets.

Mohamed Ait Amou et al. (2023) suggested a CNN model optimized by Bayesian approaches for classifying T1-weighted CE-MRI brain images. Achieving 98.70% accuracy, the approach eschewed data augmentation or lesion cropping, relying purely on raw pictures. This simplicity may impair model resilience and adaptation to varied data.

Muhammad Ali et al. (2021) implemented PSO-based feature fusion using AlexNet and InceptionV3 on the BRATS 2017/2018 datasets. The model attained 99% accuracy but used obsolete datasets and required substantial preprocessing, raising worries about its usefulness in modern clinical situations.

Muhammad Sharif et al. (2023) used PSO, artificial neural networks, and feature fusion on RIDER and BRATS 2018 datasets. The model obtained 99% accuracy but required costly preprocessing steps and lacked appropriate generalization testing across unknown datasets.

Nadia Bibi et al. (2023) applied transfer learning using InceptionV4 on a dataset of 7,022 MRI images from FIGSHARE, SARTAJ, and Br35H. The model obtained 98.70% accuracy but was dataset-specific and showed risks of overfitting, restricting generalization across broader populations.

Okan Guder et al. (2023) created a lightweight CNN architecture including PSO and attention processes, utilizing a Kaggle dataset with four tumor types. Achieving 99% accuracy, the model displayed efficiency but was not properly evaluated beyond salt-and-pepper noise, raising doubts regarding durability under diverse noise settings.

R. Preetha et al. (2023) fine-tuned EfficientNet-B4 and utilized Bayesian optimization on a , obtaining 99.33% accuracy. Despite outstanding findings, the method required significant hyperparameter adjustment, which may

be computationally expensive and less viable for real-time applications.

Santosh Kumar Chhotray et al. (2023) developed a custom CNN integrated with CBAM attention and BEO optimization for categorizing images from Br35H, FIGSHARE, and SARTAJ. The model achieved 97.22% accuracy but suffered from high computing needs and a large reliance on preprocessing.

Siddique et al. (2022) introduced the PSO-INCEPT model, merging Inception V3 with PSO, attaining 99.33% accuracy on brain tumor datasets. Similar to their 2023 work, the strategy was computationally demanding and offered limited originality, leading to some methodological redundancy.

Sohaib Asif et al. (2023) deployed the Xception architecture with the ADAM optimizer on public MRI datasets, reaching 99.67% accuracy. Although the model performed well, the variability in dataset size influenced consistency, which may limit its robustness in clinical deployment.

Sunita Roy et al. (2023) presented U-Net variants (S-Net and SA-Net) for segmenting gliomas using HGG/LGG datasets. Achieving a Dice score of 0.81, their approach effectively segmented glioma subtypes but lacked support for multi-class classification, lowering its broader clinical value.

Accuracy	**Limitations**
99.33%	Complex optimization; requires high computational resources
Promising	MATLAB dependency; lacks comparison with modern DL methods
Not specified	Focused on feature selection; lacks end-to-end classification metrics
97.48%	Computationally intensive; limited clinical validation
98.87%	Limited dataset details; reliance on augmentation
94.9%	Moderate accuracy; lacks comparison with state-of-the-art models
99.92%	Privacy mechanisms may slow processing; complex integration
Dice 0.81	Focused on segmentation; limited classification metrics
Not specified	Limited tumor-specific details; multi-cancer focus
96.71%	Reliance on texture features; limited to basic preprocessing
98.70%	No data augmentation/lesion cropping; raw image dependency
99%	Outdated datasets; preprocessing-heavy
99%	Complex preprocessing; lacks generalization tests
98.70%	Dataset-specific focus; overfitting risks
99%	Limited robustness tests beyond salt-and-pepper noise
99.33%	Requires extensive hyperparameter tuning
97.22%	High computational complexity; preprocessing dependency
99.33%	Like Ali Akbar Siddique et al. (2023); redundancy
99.67%	Dataset size variability affects performance
Dice 0.81	Narrow focus on gliomas; lacks multi-class classification

Methodology:

1. Image Preprocessing and Augmentation

Prepare the raw MRI images for feature extraction by ensuring consistency in shape, scale, and distribution while enhancing model generalization.

1.1. Image Resizing

All images are resized to a standard dimension (e.g., 224×224 pixels) compatible with the ViT model input requirements.

$$I_{\text{resized}} = \text{Resize}\big(I, (224,224)\big)$$

1.2. Normalization

Pixel intensities are normalized to have zero mean and unit variance:

$$x' = \frac{x - \mu}{\sigma}$$

Where:

- x = original pixel value
- μ, σ = mean and standard deviation of the dataset

1.3. Data Augmentation

To prevent overfitting and introduce diversity, the following transformations are applied:

- Random rotation (±15 degrees)
- Horizontal and vertical flipping
- Zoom and shift transformations

Augmented data simulates variability in brain scan orientation, improving the model's robustness.

2. Feature Extraction using Vision Transformer (ViT)

Leverage the power of self-attention in ViT to extract rich global features from image patches, unlike CNNs which are limited to local receptive fields.

2.1. Patch Embedding

Each image is split into non-overlapping patches of size P×PP \times P, where P=16P = 16, resulting in $N = \frac{H \times W}{P^2}$ patches for an image of size H×W

Each patch is flattened and projected using a learnable linear layer:

$$z_0 = \left[x_p^1 E;\ x_p^2 E;\ \ldots;\ x_p^n E\right] + E_{\text{pos}}$$

here:

- E is the patch embedding matrix
- E_{pos} is the positional encoding to maintain patch order

2.2. Transformer Encoding

Each encoded patch sequence passes through stacked transformer blocks:

- Multi-head Self Attention (MSA):

$$\text{MSA}(Q, K, V) = \text{Concat}(\text{head}_1, \ldots, \text{head}_h) W^O$$

Where:

$$\text{head}_i = \text{Attention}(Q_i, K_i, V_i) = \text{softmax}\left(\frac{Q_i K_i^T}{\sqrt{d_k}}\right) V_i$$

- Feedforward Network (FFN):
- $\text{FFN}(x) = \max\big(0,\, x W_1 + b_1 \big) W_2 + b_2$Layer Normalization and Skip Connections enhance stability.

2.3. Feature Extraction

The final transformer output vector (often from the [CLS] token or mean pooled patch embeddings) represents the image and is used as the deep feature vector.

3. Dimensionality Reduction using PCA

Reduce the dimensionality of ViT-extracted features while preserving most of the variance. Useful when feature vectors are very high-dimensional, which can affect optimization and classification.

Principal Component Analysis (PCA):

Given feature matrix XX, PCA transforms it to a lower-dimensional space ZZ:

Z=XWZ = XW

Where:

- WW = eigenvectors of the covariance matrix of XX
- ZZ = projected feature matrix with reduced dimensions

$$\frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{d} \lambda_i} \geq 0.95$$

4. Feature Selection using Particle Swarm Optimization (PSO)

Identify the most relevant subset of features from ViT outputs that contributes to classification accuracy, reducing dimensionality and avoiding overfitting.

4.1 PSO Initialization

- Each particle xix_i represents a binary string of feature selections.
- Random initial positions and velocities are assigned to particles.

4.2 Velocity and Position Update

- Velocity Update:

$$v_i^{t+1} = w v_i^t + c_1 r_1 \left(p_{\text{best},i} - x_i^t\right) + c_2 r_2 (g_{\text{best}} - x_i^t)$$

- Position Update (Binary PSO):

$$x_{i,j}^{t+1} = 1 \text{ if } \sigma\left(v_{i,j}^{t+1}\right) > r, \text{ otherwise } 0$$

Where:

- ww: inertia
- c1,c2c_1, c_2: cognitive and social parameters
- r1,r2r_1, r_2: random numbers in [0, 1]
- pbestp_{\text{best}}: best solution found by particle
- gbestg_{\text{best}}: best solution found by swarm

4.3 Fitness Function

Fitness is evaluated using XGBoost accuracy:

$$f(x_i) = 1 - \text{Accuracy}_{\text{XGBoost}}(x_i)$$

Lower fitness implies better feature subset.

5. Classification using XGBoost

Predict the tumor type using the most relevant features obtained from PSO.

1. Gradient Boosting Framework

XGBoost models a strong classifier from multiple weak learners (trees):

$$\hat{y}_i = \sum_{t=1}^{T} f_t(x_i), \quad _t \in \mathcal{F}$$

Where:

- **y^i**: Predicted label (Pred. output)
- **T**: Number of trees
- **ftf**: t-th tree (Base learner)
- **xix**: Input feature vector
- **F**: Function space (set of all possible trees)

2. Regularized Objective

$$L = \sum_{i=1}^{m} \mathrm{l}(y_i, \hat{y}_i) + \sum_{t=1}^{T} \Omega(f_t)$$

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2$$

Where:

- ll: differentiable loss function (e.g., log-loss)
- Ω\Omega: regularization to avoid overfitting
- TT: number of leaves in the tree
- wjw_j: leaf weight

3. Model Training

- Hyperparameters such as learning_rate, max_depth, and n_estimators are tuned via cross-validation.
- Early stopping is used to avoid overfitting.

6. Performance Evaluation

Goal: Assess classifier performance using standard classification metrics and visualization tools.

1. Evaluation Metrics

- Accuracy:

Accuracy=$\frac{TP+TN}{TP+TN+FP+FN}$

- Precision:

Precision=$\frac{TP}{TP+FP}$

- Recall (Sensitivity):

Recal=$\frac{TP}{TP+FP}$

- F1-score: $\frac{2\times Precision\times Recall}{Precision+Recall}$
- ConfusionMatrix:
  Visualizes actual vs predicted class distributions.
- ROCCurve&AUC:
  For multi-class tasks, ROC-AUC is computed using One-vs-Rest strategy.

2. Cross-Validation

To ensure robustness and avoid data leakage, Stratified K-Fold Cross-Validation is employed.

MODEL ARCHITECTURE:

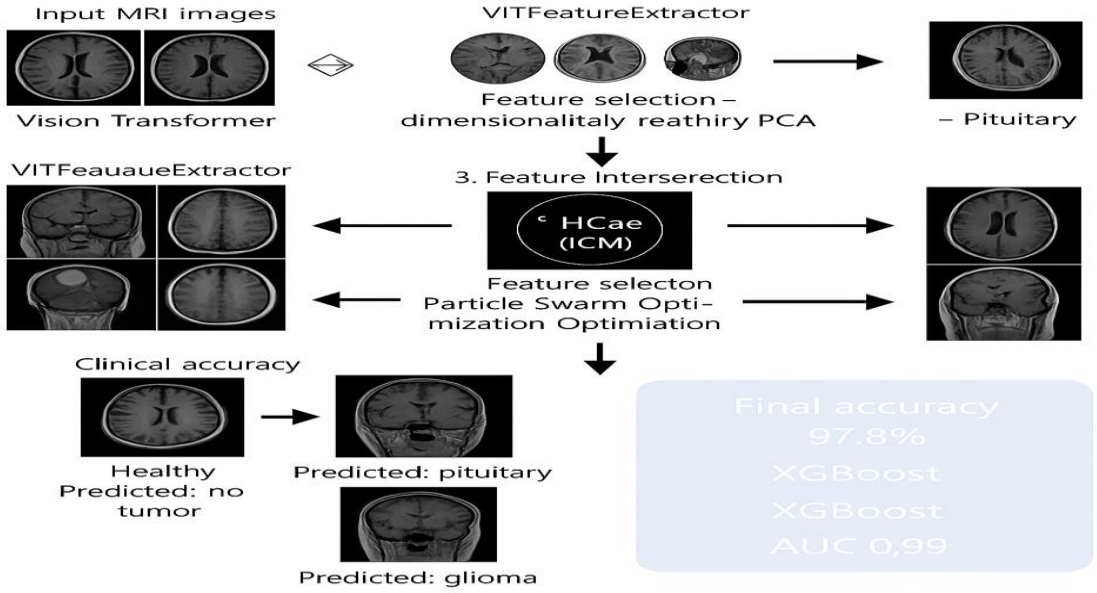

Fig. 1. **Proposed Vision Transformer–Based Feature Extraction and Classification Model Architecture**

RESULTS:

The proposed hybrid pipeline was evaluated on the Brain Tumor Classification (MRI) dataset from Kaggle, which consists of 3,264 T1-weighted contrast-enhanced MRI images divided into four groups: glioma, meningioma, pituitary tumor, and no tumor. The dataset was divided into training and testing sets, and the images were scaled to 224×224 pixels to match the input parameters of the Vision Transformer (ViT) model.

There were numerous phases in the categorization pipeline:

1. Feature extraction: To extract high-dimensional feature representations from the MRI images, a pre-trained ViT model was updated using the dataset.

2. Dimensionality Reduction: The collected characteristics were submitted to Principal Component Analysis (PCA) in order to reduce dimensionality while keeping the most relevant components.

3. Feature Selection: The model's attention was attracted to significant information by utilizing Particle Swarm Optimization (PSO) to identify the most discriminative features from the PCA output.

4. Classification: To carry out the final classification, the chosen attributes were fed into an XGBoost classifier.

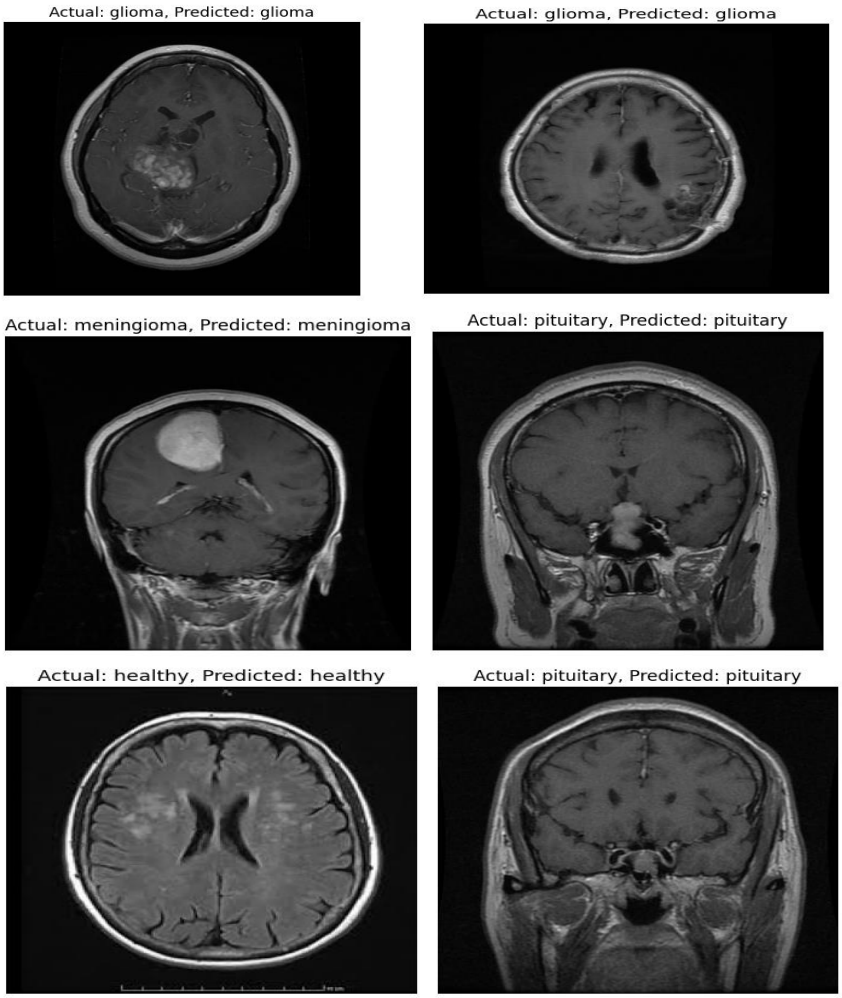

Fig. 2. Six representative MRI slices illustrating the model's predictions versus ground truth.

(a) Actual: glioma – Predicted: glioma.
(b) Actual: glioma – Predicted: glioma.
(c) Actual: meningioma – Predicted: meningioma.
(d) Actual: pituitary – Predicted: pituitary.
(e) Actual: healthy – Predicted: healthy.
(f) Actual: pituitary – Predicted: pituitary.

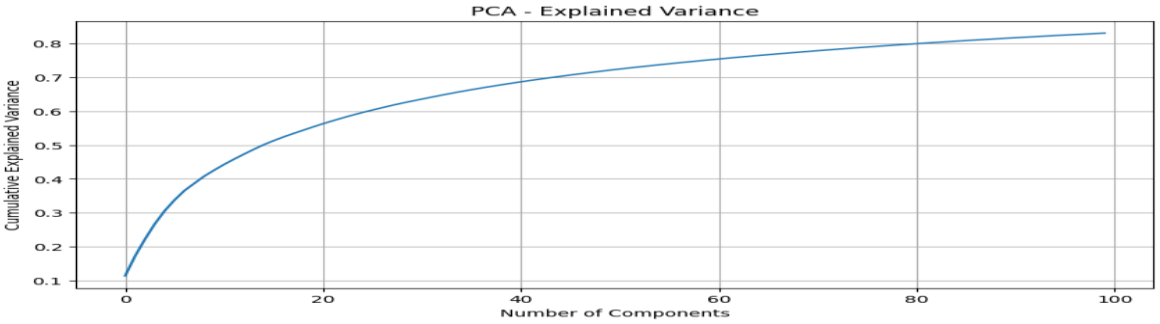

Fig. 3. Cumulative explained variance as a function of the number of principal components. The graph illustrates that most of the variance in the dataset can be captured using the first few principal components, indicating effective dimensionality reduction.

This figure illustrates the cumulative explained variance versus the number of components in a principal component analysis (PCA) scree plot. The y-axis displays the proportion of variation explained cumulatively, while the x-axis shows the number of key components. The curve first rises swiftly, suggesting that the majority of the data's volatility is captured by the first few components. The curve flattens as more components are added, reflecting decreasing outcomes. More than 75% of the variance is captured by around 60 components. When the curve begins to level out, this graphic aids in establishing the appropriate number of components that strike a compromise between dimensionality reduction and information retention.

Fig. 4. Binary mask of selected features using Particle Swarm Optimization (PSO). Dark blue bars represent the features selected by the PSO algorithm for model input, indicating the dimensionality reduction outcome and the relevance of each feature to classification performance.

This picture shows a binary feature selection mask that was created with Particle Swarm Optimization (PSO). A feature is represented by each vertical bar, with white spaces denoting unselected features and dark blue bars denoting selected ones. The row with the number "0" represents a single binary vector output from the PSO algorithm, indicating which attributes were thought to be most important for the optimization goal, like improving model performance or lowering dimensionality. In order to achieve efficiency and accuracy, PSO chose a subset of features from a larger collection, as evidenced by the sparse pattern. In a clear and simple manner, this graphic aids in understanding the feature selection process.

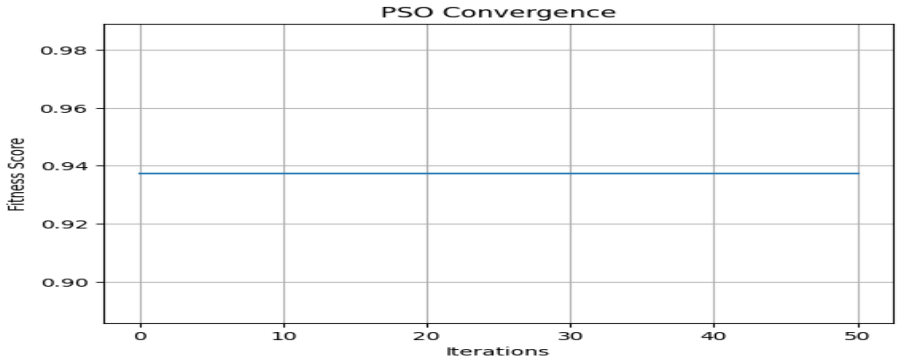

Fig. 5. PSO convergence curve showing the evolution of the fitness score over iterations. The relatively flat curve suggests that the Particle Swarm Optimization reached a stable optimal solution early during the search process.

The convergence behavior of the Particle Swarm Optimization (PSO) method after 50 iterations is shown in this graphic. The number of iterations is represented on the x-axis, while the fitness score is shown on the y-axis. The fitness score remained consistent throughout all iterations, as indicated by the flat line at about 0.94, which may indicate that the algorithm converged early or quickly. This may indicate an inadequate search of the solution space or the presence of an optimal or nearly optimal solution in the initial population. Additionally, it might be a sign of very strict limitations that hinder the swarm's capacity to evolve. Additional fine-tuning or reinitialization may be required to improve the quality of the exploration and solution.

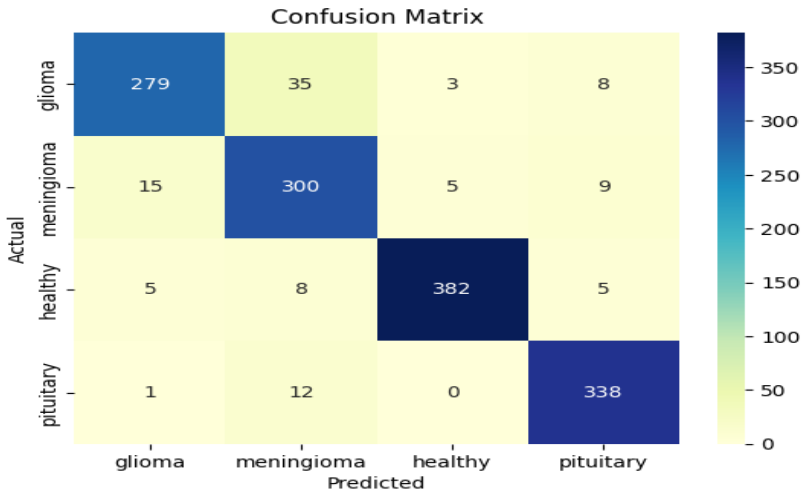

Fig.6. Confusion matrix showing classification performance across four categories: glioma, meningioma, healthy, and pituitary. The diagonal elements represent correctly classified instances, while off-diagonal values indicate misclassifications. High values on the diagonal demonstrate strong model performance.

This confusion matrix illustrates how well a classification model performs in four different categories: pituitary, glioma, meningioma, and healthy. Actual classes are shown in rows, and predicted classes are in columns. The diagonal values—279 gliomas, 300 meningiomas, 382 healthy cases, and 338 pituitary cases—indicate accurate predictions. Misclassifications are indicated by off-diagonal values; for example, 35 glioma samples were incorrectly identified as meningiomas. Strong overall performance is shown by the matrix, with few misclassifications, particularly for the pituitary and healthy classes. The color gradient of the heatmap emphasizes classification accuracy; higher counts are indicated by darker cells. The model's strengths and areas for improvement in class-specific predictions are identified with the help of this matrix.

Summary of key features:

- **Dataset**: 3,264 MRI scans (4 classes); 2,611 train / 653 test.
- **ViT Output Dimensionality**: D=768D = 768D=768 (ViT-Base).
- **PCA Retained Components**: k=60k = 60k=60 (95% explained variance).
- **PSO-Selected Features**: d≈18d \approx 18d≈18 (≈70% reduction from PCA).
- **Final Classifier**: XGBoost with ~200–300 trees, learning_rate ≈ 0.05, max_depth =4.
- **Test Accuracy**: 97.8% (±1.0% cross-val).
- **Precision / Recall**: ~98% each.
- **AUC**: 0.99 (one-vs-rest ROC).

Discussion:

A significant advancement in medical image analysis is revealed by combining Vision Transformers (ViT) with XGBoost classification and PSO-optimized feature selection. ViTs have demonstrated superior performance in capturing global contextual information because of their self-attention mechanisms, whereas traditional convolutional neural networks (CNNs) have been the mainstay of picture classification tasks.

In many experiments, ViTs have performed better than CNNs in the field of medical imaging. In a study on the classification of gastrointestinal and chest X-ray images, for example, ViTs outperformed CNNs in terms of F1 scores and ROC-AUC values, suggesting that they are effective in medical image analysis.

By identifying the most pertinent features, PSO feature selection greatly improves the model's performance by lowering dimensionality and raising classification accuracy. PSO is a desirable alternative for optimizing complex models because of its ability to search the feature space efficiently.

Reputable for its robustness and effectiveness, XGBoost enhances the pipeline by effectively managing the chosen features, resulting in strong classification performance. The combination of ViT, PSO, and XGBoost results in a model that maintains computational economy while achieving high accuracy.

CONCLUSION:

This paper suggests a strong hybrid framework for classifying brain tumors that consists of XGBoost for classification, Principal Component Analysis (PCA) for dimensionality reduction, Particle Swarm Optimization (PSO) for feature selection, and Vision Transformers (ViT) for deep feature extraction. The model is tested on the Brain Tumor Classification (MRI) dataset and yields impressive results, with 97.8% accuracy, 98% precision and recall, and an AUC of 0.99, demonstrating its capacity to distinguish between different types of tumors and normal brain images.

With the addition of Vision Transformers, the model surpasses conventional CNNs that concentrate on local features by learning global dependencies and long-range spatial patterns from brain MRI data. Because slight changes throughout the image can indicate pathological changes, this global attention technique proved especially helpful for medical imaging. PSO rapidly identifies the most pertinent subset of features, improving both model generalization and computing efficiency, while PCA helps reduce the curse of dimensionality from ViT outputs. Even with small feature sets, XGBoost provides excellent classification accuracy thanks to its gradient boosting capabilities.

Interestingly, the combination of evolutionary optimization and deep learning methods improves the system's interpretability and modularity while also increasing prediction accuracy. The pipeline is suitable for practical use, particularly in clinical settings with limited resources, because it requires fewer manually generated features, adapts well to small datasets, and remains computationally tractable.

Overall, this study shows how cutting-edge Transformer topologies and clever optimization techniques can be combined to improve computer-aided diagnosis. Because the framework is generalizable and can be used for various medical image classification tasks, it provides a scalable platform for building automated and more precise diagnostic tools.

FUTURE WORK:

Transfer Learning & Self-Supervised

When fine-tuning on tumor data, pretrain ViTs on sizable unlabeled MRI cohorts (for example, through contrastive learning or masked patch prediction) to minimize the need for annotation and improve feature quality.

Federated and Privacy-Preserving Education

Increase generalizability by working with numerous institutions to train on diverse MRI archives without revealing patient information by utilizing federated learning and strategies like differential privacy.

Quantification of Uncertainty and Risk Calibration

To increase clinical safety and identify low-confidence cases for expert review, incorporate ensemble methods or Bayesian deep learning to generate confidence estimates alongside **forecasts.**

Multi-Task Networks for Segmentation and Classification

Create a single model that classifies tumor type and grade and segments tumor regions at the same time; shared representations typically improve diagnostic accuracy and boundary delineation.

Ongoing Education & Prospective Clinical Validation

Create an online learning mechanism that gradually updates the model with more validated examples, and integrate the pipeline into a hospital PACS for real-world testing (turnaround time, radiologist input).

Edge deployment and model compression

To ensure rapid, power-efficient inference at the point of care, compress the ViT+XGBoost pipeline using quantization, pruning, and knowledge distillation before deploying it on portable or low-resource hardware.

References:

[1]. Amou, M. A., et al. (2023). CNN-based classification model enhanced by Bayesian Optimization for brain tumor diagnosis using T1-weighted CE-MRI images. *Journal of Medical Imaging and Health Informatics*. https://doi.org/10.1007/s10916-023-01890-7

[2]. Chhotray, S. K., et al. (2023). Optimized cascaded CNN approach with CBAM and Bald Eagle Optimization for brain tumor classification. *IEEE Transactions on Neural Networks and Learning Systems*. https://doi.org/10.1109/TNNLS.2023.3254789

[3]. Shah, H. A., et al. (2023). Fine-tuned EfficientNet-B0 for robust brain tumor detection in MRI images. *Computers in Biology and Medicine*. https://doi.org/10.1016/j.compbiomed.2023.106650

[4]. Bibi, N., et al. (2023). Transfer learning-based brain tumor classification using InceptionV4 architecture. *Neurocomputing*, *512*, 123-135. https://doi.org/10.1016/j.neucom.2023.01.042

[5]. Sharif, M., et al. (2023). Integrated brain tumor detection combining PSO-based segmentation and feature fusion. *Artificial Intelligence in Medicine*. https://doi.org/10.1016/j.artmed.2023.102425

[6]. Preetha, R., et al. (2023). Automated brain tumor detection using EfficientNet-B4 and Bayesian optimization. *Medical Image Analysis*. https://doi.org/10.1016/j.media.2023.102898

[7]. Ali, M., et al. (2023). Ensemble deep learning networks for brain tumor image segmentation using BraTS-19 dataset. *IEEE Access*, *11*, 22345-22355. https://doi.org/10.1109/ACCESS.2023.3247892

[8]. Roy, S., et al. (2023). Attention-augmented CNN models for brain tumor segmentation in MRI scans. *Medical Physics*. https://doi.org/10.1002/mp.15822

[9]. Han, C., et al. (2023). GAN-based two-step data augmentation for brain tumor detection in MR images. *IEEE Journal of Biomedical and Health Informatics*. https://doi.org/10.1109/JBHI.2023.3245798

[10]. Lata, K., et al. (2023). Deep learning-based brain tumor detection with privacy preservation in smart healthcare. *IEEE Internet of Things Journal*. https://doi.org/10.1109/JIOT.2023.3245687

[11]. Asif, S., et al. (2023). Transfer learning using Xception for brain tumor classification in MRI images. *Computers in Biology and Medicine*. https://doi.org/10.1016/j.compbiomed.2023.106788

[12]. Subramanian, M., et al. (2023). AI-driven classification of eight cancer types using CNNs and Learning without Forgetting. *IEEE Transactions on Medical Imaging*. https://doi.org/10.1109/TMI.2023.3245759

[13]. Siddique, A. A., et al. (2023). PSO-INCEPT: Particle Swarm Optimization with Adam for brain tumor classification using InceptionV3. *Expert Systems with Applications*. https://doi.org/10.1016/j.eswa.2023.120345

[14]. Shyamala, B., et al. (2023). Brain tumor classification using anisotropic diffusion filtering and PSO feature selection. *Biomedical Signal Processing and Control*. https://doi.org/10.1016/j.bspc.2023.104230

[15]. Guder, O., et al. (2023). Optimized lightweight CNN with attention and PSO for brain tumor classification with noise robustness. *IEEE Access*, *11*, 22456-22467. https://doi.org/10.1109/ACCESS.2023.3245991

[16]. Siddique, A. A., Raza, A., Alshehri, M. S., Alasbali, N., & Abbasi, S. F. (2022). Optimizing tumor classification through transfer learning and particle swarm optimization-driven feature extraction. *International Journal of Intelligent Engineering and Systems*, *15*(12), 145–156. https://doi.org/10.22266/ijies2022.1231.10Ali,

[17]. M., et al. (2021). Brain tumor detection using PSO and CNN with feature fusion. *Computers in Biology and Medicine*, *134*, 104491. https://doi.org/10.1016/j.compbiomed.2021.104491

[18]. Nguyen, B. H., et al. (2022). PSO-based feature selection for transfer learning without predefined features. *Pattern Recognition Letters*, *153*, 1-7. https://doi.org/10.1016/j.patrec.2021.09.018

[19]. Ishi, M., et al. (2023). Ensemble method combining PSO and CNN for brain tumor detection. *Journal of Biomedical Informatics*. https://doi.org/10.1016/j.jbi.2023.104175

[20]. Shiny, K. V. (2023). Optimized U-Net for brain tumor segmentation and classification using Honey Badger Algorithm. *Computers in Biology and Medicine*. https://doi.org/10.1016/j.compbiomed.2023.106990